Towards a Unified Approach to Memory- and Statistical-Based Machine Translation

Daniel Marcu
Information Sciences Institute and
Department of Computer Science
University of Southern California
4676 Admiralty Way, Suite 1001
Marina del Rey, CA 90292
marcu@isi.edu

Abstract

We present a set of algorithms that enable us to translate natural language sentences by exploiting both a translation memory and a statistical-based translation model. Our results show that an automatically derived translation memory can be used within a statistical framework to often find translations of higher probability than those found using solely a statistical model. The translations produced using both the translation memory and the statistical model are significantly better than translations produced by two commercial systems: our hybrid system translated perfectly 58% of the 505 sentences in a test collection, while the commercial systems translated perfectly only 40-42% of them.

1 Introduction

Over the last decade, much progress has been made in the fields of example-based (EBMT) and statistical machine translation (SMT). EBMT systems work by modifying existing, human produced translation instances, which are stored in a translation memory (TMEM). Many methods have been proposed for storing translation pairs in a TMEM, finding translation examples that are relevant for translating unseen sentences, and modifying and integrating translation fragments to produce correct outputs. Sato (1992), for example, stores complete parse trees in the TMEM and selects and generates new translations by performing similarity matchings on these trees. Veale and Way (1997) store complete sentences; new translations are generated by modifying the TMEM translation that is most similar to the input sentence. Others store phrases; new translations are produced by optimally partitioning the input into phrases that match examples from the TMEM (Maruyana and Watanabe, 1992), or by finding all partial matches and then choosing the best possible translation using a multi-engine translation system (Brown, 1999).

With a few exceptions (Wu and Wong, 1998), most SMT systems are couched in the noisy channel framework (see Figure 1). In this framework, the source language, let's say English, is assumed to be generated by a noisy probabilistic source.[1] Most of the current statistical MT systems treat this source as a sequence of words (Brown et al., 1993). (Alternative approaches exist, in which the source is taken to be, for example, a sequence of aligned templates/phrases (Wang, 1998; Och et al., 1999) or a syntactic tree (Yamada and Knight, 2001).) In the noisy-channel framework, a monolingual corpus is used to derive a statistical language model that assigns a probability to a sequence of words or phrases, thus enabling one to distinguish between sequences of words that are grammatically correct and sequences that are not. A sentence-aligned parallel corpus is then used in order to build a probabilistic translation model

[1] For the rest of this paper, we use the terms *source* and *target languages* according to the jargon specific to the noisy-channel framework. In this framework, the *source language* is the language into which the machine translation system translates.

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE **2001**	2. REPORT TYPE	3. DATES COVERED **00-00-2001 to 00-00-2001**

4. TITLE AND SUBTITLE **Towards a Unified Approach to Memory- and Statistical-Based Machine Translation**	5a. CONTRACT NUMBER
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **University of California,Information Sciences Institute ,4676 Admiralty Way,Marina del Rey,CA,90292**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE		**8**	
unclassified	**unclassified**	**unclassified**			

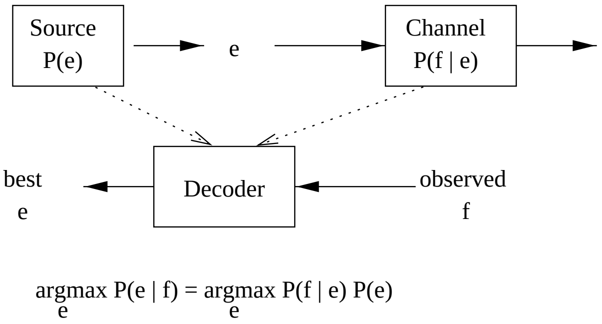

Figure 1: The noisy channel model.

that explains how the source can be turned into the target and that assigns a probability to every way in which a source e can be mapped into a target f. Once the parameters of the language and translation models are estimated using traditional maximum likelihood and EM techniques (Dempster et al., 1977), one can take as input any string in the target language f, and find the source e of highest probability that could have generated the target, a process called decoding (see Figure 1).

It is clear that EBMT and SMT systems have different strengths and weaknesses. If a sentence to be translated or a very similar one can be found in the TMEM, an EBMT system has a good chance of producing a good translation. However, if the sentence to be translated has no close matches in the TMEM, then an EBMT system is less likely to succeed. In contrast, an SMT system may be able to produce perfect translations even when the sentence given as input does not resemble any sentence from the training corpus. However, such a system may be unable to generate translations that use idioms and phrases that reflect long-distance dependencies and contexts, which are usually not captured by current translation models.

This paper advances the state-of-the-art in two respects. First, we show how one can use an existing statistical translation model (Brown et al., 1993) in order to automatically derive a statistical TMEM. Second, we adapt a decoding algorithm so that it can exploit information specific both to the statistical TMEM and the translation model. Our experiments show that the automatically derived translation memory can be used within the statistical framework to often find translations of higher probability than those found using solely the statistical model. The translations produced using both the translation memory and the statistical model are significantly better than translations produced by two commercial systems.

2 The IBM Model 4

For the work described in this paper we used a modified version of the statistical machine translation tool developed in the context of the 1999 Johns Hopkins' Summer Workshop (Al-Onaizan et al., 1999), which implements IBM translation model 4 (Brown et al., 1993).

IBM model 4 revolves around the notion of word alignment over a pair of sentences (see Figure 2). The word alignment is a graphical representation of an hypothetical stochastic process by which a source string e is converted into a target string f. The probability of a given alignment a and target sentence f given a source sentence e is given by

$$
\begin{aligned}
&\mathrm{P(a, f \mid e)} = \\
&\prod_{i=1}^{l} \mathrm{n}(\phi_i|\mathrm{e}_i) \times \prod_{i=1}^{l}\prod_{k=1}^{\phi_i} \mathrm{t}(\tau_{ik}|\mathrm{e}_i)\times \\
&\prod_{i=1,\phi_i>0}^{l} \mathrm{d}_1(\pi_{i1} - c_{\rho_i}|class(\mathrm{e}_{\rho_i}), class(\tau_{i1}))\times \\
&\prod_{i=1}^{l}\prod_{k=2}^{\phi_i} \mathrm{d}_{>1}(\pi_{ik} - \pi_{i(k-1)}|class(\tau_{ik}))\times \\
&\begin{pmatrix} m-\phi_0 \\ \phi_0 \end{pmatrix} {p_1}^{\phi_0}(1-p_1)^{m-2\phi_0}\times \\
&\prod_{k=1}^{\phi_0} t(\tau_{0k}|\mathrm{NULL})
\end{aligned}
$$

where the factors delineated by × symbols correspond to hypothetical steps in the following generative process:

- Each English word e_i is assigned with probability $\mathrm{n}(\phi_i|\mathrm{e}_i)$ a fertility ϕ_i, which corresponds to the number of French words into which e is going to be translated.

- Each English word e_i is then translated with probability $\mathrm{t}(\tau_{ik}|\mathrm{e}_i)$ into a French word τ_{ik}, where k ranges from 1 to the number of words ϕ_i (fertility of e_i) into which e_i is translated. For example, the English word "no" in Figure 2 is a word of fertility 2 that is translated into "aucun" and "ne".

- The rest of the factors denote distorsion probabilities (d), which capture the probability that words change their position when translated from one language into another; the probability of some French words being generated from an invisible English NULL element (p_1), etc. See (Brown et al., 1993) or (Germann et al., 2001) for a detailed discussion of this translation model and a description of its parameters.

3 Building a statistical translation memory

Companies that specialize in producing high-quality human translations of documentation and news rely often on translation memory tools to increase their productivity (Sprung, 2000). Building high-quality TMEM is an expensive process that requires many person-years of work. Since we are not in the fortunate position of having access to an existing TMEM, we decided to build one automatically.

We trained IBM translation model 4 on 500,000 English-French sentence pairs from the Hansard corpus. We then used the Viterbi alignment of each sentence, i.e., the alignment of highest probability, to extract tuples of the form $\langle e_i, e_{i+1}, \ldots, e_{i+k}; f_j, f_{j+1}, \ldots, f_{j+l}; a_j, a_{j+1}, \ldots, a_{j+l}\rangle$, where $e_i, e_{i+1}, \ldots, e_{i+k}$ represents a contiguous English phrase, $f_j, f_{j+1}, \ldots, f_{j+l}$ represents a contiguous French phrase, and $a_j, a_{j+1}, \ldots, a_{j+l}$ represents the Viterbi alignment between the two phrases. We selected only "contiguous" alignments, i.e., alignments in which the words in the English phrase generated only words in the French phrase and each word in the French phrase was generated either by the NULL word or a word from the English phrase. We extracted only tuples in which the English and French phrases contained at least two words.

For example, in the Viterbi alignment of the two sentences in Figure 2, which was produced automatically, "there" and "." are words of fertility 0, NULL generates the French lexeme ".", "is" generates "est", "no" generates "aucun" and "ne", and so on. From this alignment we extracted the six tuples shown in Table 1, because they were the only ones that satisfied all conditions mentioned above. For example, the pair ⟨ no one ; aucun syndicat particulier ne ⟩ does not occur in the translation memory because the French word "syndicat" is generated by the word "union", which does not occur in the English phrase "no one".

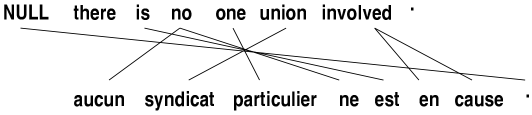

Figure 2: Example of Viterbi alignment produced by IBM model 4.

By extracting all tuples of the form $\langle e; f; a\rangle$ from the training corpus, we ended up with many duplicates and with French phrases that were paired with multiple English translations. We chose for each French phrase only one possible English translation equivalent. We tried out two distinct methods for choosing a translation equivalent, thus constructing two different probabilistic TMEMs:

- The Frequency-based Translation MEMory (FTMEM) was created by associating with each French phrase the English equivalent that occurred most often in the collection of phrases that we extracted.

- The Probability-based Translation MEMory (PTMEM) was created by associating with each French phrase the English equivalent that corresponded to the alignment of highest probability.

In contrast to other TMEMs, our TMEMs explicitly encode not only the mutual translation pairs but also their corresponding word-level alignments, which are derived according to a certain translation model (in our case, IBM model 4). The mutual translations can be anywhere between two words long to complete sentences. Both methods yielded translation memories that contained around 11.8 million word-aligned translation pairs. Due to efficiency considerations and memory limitations — the software we wrote loads a complete TMEM into the memory — we used in our experiments only a fraction of the TMEMs, those that contained phrases at most 10

English	French	Alignment
one union	syndicat particulier	one → {particulier}; union→ {syndicat}

no one union	aucun syndicat particulier ne	no → {aucun, ne};
		one → {particulier}; union→ {syndicat}

is no one union	aucun syndicat particulier ne est	is → {est}; no → {aucun, ne};
		one → {particulier}; union→ {syndicat}

there is no one union	aucun syndicat particulier ne est	is → {est}; no → {aucun, ne};
		one → {particulier}; union→ {syndicat}

is no one union involved	aucun syndicat particulier ne est en cause	is → {est}; no → {aucun, ne};
		one → {particulier}; union→ {syndicat}
		involved → {en cause}

there is no one union involved	aucun syndicat particulier ne est en cause	is → {est}; no → {aucun, ne};
		one → {particulier}; union→ {syndicat}
		involved → {en cause}

there is no one union involved .	aucun syndicat particulier ne est en cause .	is → {est}; no → {aucun, ne};
		one → {particulier}; union→ {syndicat}
		involved → {en cause}; NULL → { . }

Table 1: Examples of automatically constructed statistical translation memory entries.

TMEM	Perfect	Almost perfect	Incorrect	Unable to judge
FTMEM	62.5%	8.5%	27.0%	2.0%
PTMEM	57.5%	7.5%	33.5%	1.5%

Table 2: Accuracy of automatically constructed TMEMs.

words long. This yielded a working FTMEM of 4.1 million and a PTMEM of 5.7 million phrase translation pairs aligned at the word level using IBM statistical model 4.

To evaluate the quality of both TMEMs we built, we extracted randomly 200 phrase pairs from each TMEM. These phrases were judged by a bilingual speaker as

- *perfect* translations if she could imagine contexts in which the aligned phrases could be mutual translations of each other;
- *almost perfect* translations if the aligned phrases were mutual translations of each other and one phrase contained one single word with no equivalent in the other language[2];
- *incorrect* translations if the judge could not imagine any contexts in which the aligned phrases could be mutual translations of each other.

The results of the evaluation are shown in Table 2. A visual inspection of the phrases in our TMEMs and the judgments made by the evaluator suggest that many of the translations labeled as incorrect make sense when assessed in a larger context. For example, “autres régions de le pays que” and “other parts of Canada than” were judged as incorrect. However, when considered in a context in which it is clear that “Canada” and “pays” corefer, it would be reasonable to assume that the translation is correct. Table 3 shows a few examples of phrases from our FTMEM and their corresponding correctness judgments.

Although we found our evaluation to be extremely conservative, we decided nevertheless to stick to it as it adequately reflects constraints specific to high-standard translation environments in which TMEMs are built manually and constantly checked for quality by specialized teams (Sprung, 2000).

4 Statistical decoding using both a statistical TMEM and a statistical translation model

The results in Table 2 show that about 70% of the entries in our translation memory are correct or almost correct (very easy to fix). It is, though, an empirical question to what extend such TMEMs can be used to improve the performance of current translation systems. To determine this, we modified an existing decoding algorithm so that it can exploit information specific both to a statistical translation model and a statistical TMEM.

[2] For example, the translation pair “final , le secrétaire de” and “final act , the secretary of” were labeled as almost perfect because the English word “act” has no French equivalent.

English	French	Judgment
, but I cannot say	, mais je ne puis dire	correct
how did this all come about ?	comment est-ce arrivée ?	correct
but , I humbly believe	mais , à mon humble avis	correct
final act , the secretary of	final , le secrétaire de	almost correct
other parts of Canada than	autres régions de le pays que	incorrect
what is the total amount accumulated	a combien se élève la	incorrect
that party present this	ce parti présent aujourd'hui	incorrect
the airraft company to present further studies	de autre études	incorrect

Table 3: Examples of TMEM entries with correctness judgments.

The decoding algorithm that we use is a greedy one — see (Germann et al., 2001) for details. The decoder guesses first an English translation for the French sentence given as input and then attempts to improve it by exploring greedily alternative translations from the immediate translation space. We modified the greedy decoder described by Germann et al. (2001) so that it attempts to find good translation starting from two distinct points in the space of possible translations: one point corresponds to a word-for-word "gloss" of the French input; the other point corresponds to a translation that resembles most closely translations stored in the TMEM.

As discussed by Germann et al. (2001), the word-for-word gloss is constructed by aligning each French word f_j with its most likely English translation $\mathrm{e}_{\mathrm{f}_j}$ ($\mathrm{e}_{\mathrm{f}_j} = \mathrm{argmax}_e\ \mathrm{t}(\mathrm{e} \mid \mathrm{f}_j)$). For example, in translating the French sentence "Bien entendu , il parle de une belle victoire .", the greedy decoder initially assumes that a good translation of it is "Well heard , it talking a beautiful victory" because the best translation of "bien" is "well", the best translation of "entendu" is "heard", and so on. A word-for-word gloss results (at best) in English words written in French word order.

The translation that resembles most closely translations stored in the TMEM is constructed by deriving a "cover" for the input sentence using phrases from the TMEM. The derivation attempts to cover with translation pairs from the TMEM as much of the input sentence as possible, using the longest phrases in the TMEM. The words in the input that are not part of any phrase extracted from the TMEM are glossed. For example, this approach may start the translation process from the phrase "well , he is talking a beautiful victory" if the TMEM contains the pairs ⟨well , ; bien entendu ,⟩ and ⟨he is talking; il parle⟩ but no pair with the French phrase "belle victoire".

If the input sentence is found "as is" in the translation memory, its translation is simply returned and there is no further processing. Otherwise, once an initial alignment is created, the greedy decoder tries to improve it, i.e., it tries to find an alignment (and implicitly a translation) of higher probability by modifying locally the initial alignment. The decoder attempts to find alignments and translations of higher probability by employing a set of simple operations, such as changing the translation of one or two words in the alignment under consideration, inserting into or deleting from the alignment words of fertility zero, and swapping words or segments.

In a stepwise fashion, starting from the initial gloss or initial cover, the greedy decoder iterates exhaustively over all alignments that are one such simple operation away from the alignment under consideration. At every step, the decoder chooses the alignment of highest probability, until the probability of the current alignment can no longer be improved.

5 Evaluation

We extracted from the test corpus a collection of 505 French sentences, uniformly distributed across the lengths 6, 7, 8, 9, and 10. For each French sentence, we had access to the human-generated English translation in the test corpus, and to translations generated by two commercial systems. We produced translations using three versions of the greedy decoder: one used only the statistical translation model, one used the translation model and the FTMEM, and one used the translation model and the PTMEM.

We initially assessed how often the translations obtained from TMEM seeds had higher proba-

Sent. length	Found in FTMEM	Higher prob. from FTMEM	Same result	Higher prob. from gloss
6	33	9	43	16
7	27	9	48	17
8	29	16	42	14
9	31	15	28	27
10	31	9	43	18
All (%)	30%	12%	40%	18%

Table 4: The utility of the FTMEM.

Sent. length	Found in FTMEM	Higher prob. from FTMEM	Same result	Higher prob. from gloss
6	33	9	43	16
7	27	10	50	14
8	30	16	41	14
9	31	15	36	19
10	31	15	31	13
All (%)	31%	13%	41%	15%

Table 5: The utility of the PTMEM.

bility than the translations obtained from simple glosses. Tables 4 and 5 show that the translation memories significantly help the decoder find translations of high probability. In about 30% of the cases, the translations are simply copied from a TMEM and in about 13% of the cases the translations obtained from a TMEM seed have higher probability that the best translations obtained from a simple gloss. In 40% of the cases both seeds (the TMEM and the gloss) yield the same translation. Only in about 15-18% of the cases the translations obtained from the gloss are better than the translations obtained from the TMEM seeds. It appears that both TMEMs help the decoder find translations of higher probability consistently, across all sentence lengths.

In a second experiment, a bilingual judge scored the human translations extracted from the automatically aligned test corpus; the translations produced by a greedy decoder that use both TMEM and gloss seeds; the translations produced by a greedy decoder that uses only the statistical model and the gloss seed; and translations produced by two commercial systems (A and B).

- If an English translation had the very same meaning as the French original, it was considered semantically correct. If the meaning was just a little different, the translation was considered semantically incorrect. For example, “this is rather provision disturbing” was judged as a correct semantical translation of “voilà une disposition plotôt inquiétante”, but “this disposal is rather disturbing” was judged as incorrect.
- If a translation was perfect from a grammatical perspective, it was considered to be grammatical. Otherwise, it was considered incorrect. For example, “this is rather provision disturbing” was judged as ungrammatical, although one may very easily make sense of it.

We decided to use such harsh evaluation criteria because, in previous experiments, we repeatedly found that harsh criteria can be applied consistently. To ensure consistency during evaluation, the judge used a specialized interface: once the correctness of a translation produced by a system S was judged, the same judgment was automatically recorded with respect to the other systems as well. This way, it became impossible for a translation to be judged as correct when produced by one system and incorrect when produced by another system.

Table 6, which summarizes the results, displays the percent of perfect translations (both semantically and grammatically) produced by a variety of systems. Table 6 shows that translations produced using both TMEM and gloss seeds are much better than translations that do not use TMEMs. The translation systems that use both a TMEM and the statistical model outperform significantly the two commercial systems. The figures in Table 6 also reflect the harshness of our evaluation metric: only 82% of the human translations extracted from the test corpus were considered perfect translation. A few of the errors were genuine, and could be explained by failures of the sentence alignment program that was used to create the corpus (Melamed, 1999). Most of the errors were judged as semantic, reflecting directly the harshness of our evaluation metric.

6 Discussion

The approach to translation described in this paper is quite general. It can be applied in conjunction with other statistical translation mod-

Sentence length	Humans	Greedy with FTMEM	Greedy with PTMEM	Greedy without TMEM	Commercial system A	Commercial system B
6	92	72	70	52	55	59
7	73	58	52	37	42	43
8	80	53	52	30	38	29
9	84	53	53	37	40	35
10	85	57	60	36	40	37
All(%)	82%	58%	57%	38%	42%	40%

Table 6: Percent of perfect translations produced by various translation systems and algorithms.

els. And it can be applied in conjunction with existing translation memories. To do this, one would simply have to train the statistical model on the translation memory provided as input, determine the Viterbi alignments, and enhance the existing translation memory with word-level alignments as produced by the statistical translation model. We suspect that using manually produced TMEMs can only increase the performance as such TMEMs undergo periodic checks for quality assurance.

The work that comes closest to using a statistical TMEM similar to the one we propose here is that of Vogel and Ney (2000), who automatically derive from a parallel corpus a hierarchical TMEM. The hierarchical TMEM consists of a set of transducers that encode a simple grammar. The transducers are automatically constructed: they reflect common patterns of usage at levels of abstractions that are higher than the words. Vogel and Ney (2000) do not evaluate their TMEM-based system, so it is difficult to empirically compare their approach with ours. From a theoretical perspective, it appears though that the two approaches are complementary: Vogel and Ney (2000) identify abstract patterns of usage and then use them during translation. This may address the data sparseness problem that is characteristic to any statistical modeling effort and produce better translation parameters.

In contrast, our approach attempts to stir the statistical decoding process into directions that are difficult to reach when one relies only on the parameters of a particular translation model. For example, the two phrases “il est mort” and “he kicked the bucket” may appear only in one sentence in an arbitrary large corpus. The parameters learned from the entire corpus will very likely associate very low probability to the words “kicked” and “bucket” being translated into “est” and “mort”. Because of this, a statistical-based MT system will have trouble producing a translation that uses the phrase “kick the bucket”, no matter what decoding technique it employs. However, if the two phrases are stored in the TMEM, producing such a translation becomes feasible.

If optimal decoding algorithms capable of searching exhaustively the space of all possible translations existed, using TMEMs in the style presented in this paper would never improve the performance of a system. Our approach works because it biases the decoder to search in subspaces that are likely to yield translations of high probability, subspaces which otherwise may not be explored. The bias introduced by TMEMs is a practical alternative to finding optimal translations, which is NP-complete (Knight, 1999).

It is clear that one of the main strengths of the TMEM is its ability to encode contextual, long-distance dependencies that are incongruous with the parameters learned by current context poor, reductionist channel models. Unfortunately, the criterion used by the decoder in order to choose between a translation produced starting from a gloss and one produced starting from a TMEM is biased in favor of the gloss-based translation. It is possible for the decoder to produce a perfect translation using phrases from the TMEM, and yet, to discard the perfect translation in favor of an incorrect translation of higher probability that was obtained from a gloss (or from the TMEM). It would be desirable to develop alternative ranking techniques that would permit one to prefer in some instances a TMEM-based translation, even though that translation is not the best according to the probabilistic channel model. The examples in Table 7 shows though that this is not trivial: it is not always the case that the translation of high-

Translations	Does this translation use TMEM phrases?	Is this translation correct?	Is this the translation of highest probability?
monsieur le président , je aimerais savoir .			
mr. speaker , i would like to know .	yes	yes	yes
mr. speaker , i would like to know .	no	yes	yes
je ne peux vous entendre , brian .			
i cannot hear you , brian .	yes	yes	yes
i can you listen , brian .	no	no	no
alors , je termine là - dessus .			
therefore , i will conclude my remarks .	yes	yes	no
therefore , i conclude - over .	no	no	yes

Table 7: Example of system outputs, obtained with or without TMEM help.

est probability is the perfect one. The first French sentence in Table 7 is correctly translated with or without help from the translation memory. The second sentence is correctly translated only when the system uses a TMEM seed; and fortunately, the translation of highest probability is the one obtained using the TMEM seed. The translation obtained from the TMEM seed is also correct for the third sentence. But unfortunately, in this case, the TMEM-based translation is not the most probable.

Acknowledgments. This work was supported by DARPA-ITO grant N66001-00-1-9814.

References

Yaser Al-Onaizan, Jan Curin, Michael Jahr, Kevin Knight, John Lafferty, Dan Melamed, Franz-Josef Och, David Purdy, Noah A. Smith, and David Yarowsky. 1999. Statistical machine translation. Final Report, JHU Summer Workshop.

Peter F. Brown, Stephen A. Della Pietra, Vincent J. Della Pietra, and Robert L. Mercer. 1993. The mathematics of statistical machine translation: Parameter estimation. *Computational Linguistics*, 19(2):263–311.

Ralph D. Brown. 1999. Adding linguistic knowledge to a lexical example-based translation system. In *Proceedings of TMI'99*, pages 22–32, Chester, England.

A. P. Dempster, N. M. Laird, and D. B. Rubin. 1977. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society*, 39(Ser B):1–38.

Ulrich Germann, Mike Jahr, Kevin Knight, Daniel Marcu, and Kenji Yamada. 2001. Fast decoding and optimal decoding for machine translation. In *Proceedings of ACL'01*, Toulouse, France.

Kevin Knight. 1999. Decoding complexity in word-replacement translation models. *Computational Linguistics*, 25(4).

H. Maruyana and H. Watanabe. 1992. Tree cover search algorithm for example-based translation. In *Proceedings of TMI'92*, pages 173–184.

Dan Melamed. 1999. Bitext maps and alignment via pattern recognition. *Computational Linguistics*, 25(1):107–130.

Franz Josef Och, Christoph Tillmann, and Herman Ney. 1999. Improved alignment models for statistical machine translation. In *Proceedings of the EMNLP and VLC*, pages 20–28, University of Maryland, Maryland.

S. Sato. 1992. CTM: an example-based translation aid system using the character-based match retrieval method. In *Proceedings of the 14th International Conference on Computational Linguistics (COLING'92)*, Nantes, France.

Robert C. Sprung, editor. 2000. *Translating Into Success: Cutting-Edge Strategies For Going Multilingual In A Global Age*. John Benjamins Publishers.

Tony Veale and Andy Way. 1997. Gaijin: A template-based bootstrapping approach to example-based machine translation. In *Proceedings of "New Methods in Natural Language Processing"*, Sofia, Bulgaria.

S. Vogel and Herman Ney. 2000. Construction of a hierarchical translation memory. In *Proceedings of COLING'00*, pages 1131–1135, Saarbrücken, Germany.

Ye-Yi Wang. 1998. *Grammar Inference and Statistical Machine Translation*. Ph.D. thesis, Carnegie Mellon University. Also available as CMU-LTI Technical Report 98-160.

Dekai Wu and Hongsing Wong. 1998. Machine translation with a stochastic grammatical channel. In *Proceedings of ACL'98*, pages 1408–1414, Montreal, Canada.

Kenji Yamada and Kevin Knight. 2001. A syntax-based statistical translation model. In *Proceedings of ACL'01*, Toulouse, France.